2000

Separating Auxiliary Arity Hierarchy of First-Order Incremental Evaluation Using (3+1)-ary Input Relations

Guozhu Dong
Wright State University - Main Campus, guozhu.dong@wright.edu

Louxin Zhang

Repository Citation

Dong, G., & Zhang, L. (2000). Separating Auxiliary Arity Hierarchy of First-Order Incremental Evaluation Using (3+1)-ary Input Relations. *International Journal of Foundations of Computer Science, 11* (4), 573-578.
https://corescholar.libraries.wright.edu/knoesis/405

Separating Auxiliary Arity Hierarchy of First-Order Incremental Evaluation Using (3k+1)-ary Input Relations *

Guozhu Dong
Univ. of Melbourne

Louxin Zhang
Inst. of Systems Science

4 August 1997

Key words: Arity hierarchy, on-line evaluation, incremental evaluation, database view maintenance, first-order logic, dynamic inductive definition, help bit, auxiliary relation.

1 The Main Result

A first-order incremental evaluation system (foies) uses first-order queries to maintain a database view which is defined by a non-first-order query; some auxiliary relations (views) may also need to be maintained similarly.

In [7] it was shown that using $(k-1)$-ary auxiliary relations is strictly less powerful than using k-ary ones, using a reduction to a result by Cai [3] on help bits. However, that separation was achieved using queries having input relations of arity at least $6k$. In this note we improve the separation result by reducing the arity of the input relations of the queries to $3k+1$. (Some necessary background knowledge is given in the next section.)

*This work was supported in part by a research grant from the Australian Research Council.
Part of work by the second author was done while visiting University of Melbourne.
Guozhu Dong's address: Department of Computer Science, University of Melbourne, Parkville, Vic 3052, Australia. Email: dong@cs.mu.oz.au.
Louxin Zhang's address: Institute of Systems Science, Heng Mui Keng Terrace, Singapore 119597. E-mail: lxzhang@iss.nus.sg.

Theorem 1 For each integer $k > 0$, $\mathrm{FOIES}_{k-1} \subseteq\neq \mathrm{FOIES}_k$. In fact, there is a query, over input relations of arity $3k+1$, in $\mathrm{FOIES}_k - \mathrm{FOIES}_{k-1}$.

As will be discussed in the next section, this result can be viewed as one step towards answering the question of whether graph problems such as transitive closure of directed graphs have foies, or whether some PTIME-complete graph problems have foies. By reducing the arity of queries needed in separating FOIES_k and FOIES_{k-1}, we hope to contribute towards the understanding of the power of foies. A further research problem is to improve the result of this paper by reducing $3k+1$ to $k+1$, or k, or even a constant.

We will prove this by modifying Cai's result [3] and by modifying the reduction used in [7].

Section 2 provides a brief review the notion of "first-order incremental evaluation systems". Section 3 establishes a necessary technical lemma, which is a variant of Cai's theorem. Section 4 gives the proof of the above theorem.

2 First-Order Incremental Evaluation Systems

It is common knowledge that many useful database queries such as transitive closure of (un)directed graphs and parity (whether the number of tuples in a relation is even) cannot be expressed in first-order logic [1]. Interestingly, many materialized database views defined by such queries Q can be maintained using first-order queries [9, 10, 8, 5, 6, 7].

Roughly speaking, such maintenance for the view defined by Q is carried out through a set of first-order queries (fixed for Q), which is called a first-order incremental evaluation system (or foies for short)[1]. One of these queries directly maintains the answer to Q, while the others maintain some auxiliary relations (views); they are used to derive the new views (Q or auxiliary) after each permissible update[2] to the database. Thus each of these first-order queries has as input the old database, the old answer view, the old auxiliary views, and the update.

Being able to maintain non first-order views using first-order queries is desirable for two main reasons: (i) Such maintenance can be implemented in all relational database systems, since first-order queries are available in every relational database system, even though the views themselves cannot be defined in first order. (ii) First-order maintenance algorithms

[1]Patnaik and Immerman's DynFO [10] is very similar, though different from our foies.

[2]Permissible updates are those updates whose sizes are bounded by a constant dependent only on the query Q and which transforms the old database in the domain of Q to a new database in the domain of Q.

have great potential to be adapted for parallel execution, since they have constant parallel complexity [1]. Such maintenance may also be of interest from a mathematical logic and descriptive complexity perspective.

There have been attempts to ascertain these two research problems: whether the transitive closure of arbitrary directed graphs can be maintained in first-order after edge deletions and whether the same generation query can be maintained after edge insertions. Clearly, these problems can be settled in the positive by finding first-order maintenance queries which only use auxiliary relations of arity k for some k; and can be settled in the negative by showing that there is no such k.

It is thus interesting to understand the power of foies using auxiliary relations of fixed arities. Starting from [6], the maximum arity of the auxiliary relations has been used as a measure of how hard it is to maintain a query using foies. Observe that, with maximal arity k, the auxiliary relations can hold at most $O(n^k)$ tuples, where n is the number of constants in the input database.

For each natural number k, let FOIES_k denote the class of queries having foies using k-ary auxiliary relations. Obviously, $\text{FOIES}_{k-1} \subseteq \text{FOIES}_k \; \forall k > 0$. In [6, 7] it was shown that $\text{FOIES}_{k-1} \subset \text{FOIES}_k$ for all $k > 0$. The separation for the small arities were achieved using queries whose input relations are unary or binary. However, the separation of $\text{FOIES}_{k-1} \subset \text{FOIES}_k$, $k \geq 3$, were achieved through queries whose input relations are $6k$-ary.

We now briefly review some previous results on the maintenance of the transitive closure of graphs of various kinds. In [9, 8] some foies using binary auxiliary relations (for insertion only) were given for generalized transitive closure of labelled graphs. For the transitive closure of acyclic directed graphs, [4, 5] gave a foies with no auxiliary relations. For undirected graphs, there is a foies [10] using ternary auxiliary relations; it maintains an undirected spanning forest for the undirected graphs, from which the reachability relation can be extracted. There is also a foies [6] using binary auxiliary relations; it maintains a directed spanning forest of the undirected graphs, plus some approximation of a total order on the nodes in the graph; it was shown [6] that there is no foies using unary auxiliary relations for transitive closure of undirected graphs.

3 Lemma on Help Bits

For indeterminate sets X and B, define

$$\begin{aligned} \mathcal{R} &= Z_3[X]/(x^2 - x \mid x \in X), \text{and} \\ \mathcal{R}^B &= Z_3[X \cup B]/(x^2 - x,\ b^2 - b \mid x \in X,\ b \in B). \end{aligned}$$

Lemma 2 (1). $\mathcal{R}$ is a commutative algebra of dimension $2^{|X|}$.

(2). For any ideal $\mathcal{K}$ in $\mathcal{R}$, $dim(\mathcal{R}/\mathcal{K}) + dim(\mathcal{K}) = 2^{|X|}$.

Let $\mathcal{F}$ be the class of Boolean functions on X. We have an injection ϕ from $\mathcal{F}$ to $\mathcal{R}$ satisfying:

$$\begin{aligned}
&\phi(0) = 0, \phi(1) = 1, \phi(x) = x,\\
&\phi(f \vee g) = \phi(f) + \phi(g) - \phi(f) \cdot \phi(g),\\
&\phi(f \wedge g) = \phi(f) \cdot \phi(g),\\
&\phi(\neg f) = 1 - \phi(f).
\end{aligned}$$

Thus, each Boolean function takes a form in $\mathcal{R}$.

Consider a multi-output Boolean circuit C of $\vee$, $\wedge$ and $\neg$ gates, where $\vee$ and $\wedge$ gates have unbounded fan-in. We will use $|C|$ to denote the number of gates in C. Assume C has input variables $X \cup B$, and m outputs $f_1, f_2, \cdots, f_m$, where $B = \{b_1, b_2, \cdots, b_s\}$ will be considered as help bits. The following lemma is a consequence of a result of Razborov [11] (see also [2, 12] for its generalization), which appeared in [3].

Lemma 3 Let $e, d > 0$ be real numbers and C any circuit of depth k with $|C| \leq e3^{d^{1/k}/2}$. Let h_i $(1 \leq i \leq s)$ be any Boolean functions, and f_i $(1 \leq i \leq m)$ be the outputs of C when substituting $h_i(X)$ for b_i, $1 \leq i \leq s$. Then there exist an ideal $\mathcal{K}$ of $dim(\mathcal{K}) \leq 2^{|X|}e$, and polynomials $p_i(X, B) \in \mathcal{R}^B$ of degree $\leq d$, such that $f_i(X) = p_i(X, h_1(X), \cdots, h_s(X))$ is in $\mathcal{R}/\mathcal{K}$.

Now we consider m parity functions $\Pi_i = x_{i1} \oplus \cdots \oplus x_{in}$, $1 \leq i \leq m$. Let $N = mn$ and

$$X = \{x_{ij} | 1 \leq i \leq m, 1 \leq j \leq n\}.$$

In the algebra $\mathcal{R}$, each parity function Π_i has the form of $\prod_j (1 + x_{ij}) - 1$. The following result is a variant of theorem 3.2 in [3], replacing Cai's $m \leq (mn)^{1/5-\delta}$ by $m^{2+5\delta} \leq n$, and has a similar proof. For completeness, we give a full proof.

Lemma 4 Let $\delta > 0$ and $m^{2+5\delta} \leq n$. Suppose a Boolean unbounded fan-in circuit C of depth k computes all m parity functions $\Pi_i = x_{i1} \oplus x_{i2} \oplus \cdots \oplus x_{in}$ with the help of $m - \log m - 1$ functions $h_i(X)$, $1 \leq i \leq m - \log m - 1$. Then $\exists \lambda = \lambda_{\delta,k}$ such that $|C| = \Omega(2^{N^\lambda})$.

Proof. Let $d = N^\delta$ and $c = 1 + \log m$. Suppose $|C| \leq e(N)3^{N^{\delta/k}/2}$, where $e(N)$ will be determined later. Then we have m polynomials $p_i(X, B) \in \mathcal{R}^B$ of degree $\leq N^\delta$, and an ideal

$\mathcal{K}$ of dimension $\leq e(N)2^N$, such that $\Pi_i(X) = p_i(X, h_1(X), h_2(X), \cdots, h_{m-c}(X))$ is in $\mathcal{R}/\mathcal{K}$ for $i = 1, \cdots, m$.

We observe that each h_i is $\{0, 1\}$-valued, thus $h_i^2 = h_i$ for all i. Hence, there is a canonical representation for each p_i

$$p_i = \sum_{S \subseteq [1, m-c]} f_i^S \cdot \prod_{j \in S} h_j(X),$$

where $deg(f_i^S) \leq N^\delta$ and $[1, m-c]$ denotes the set of integers between 1 and $m - c$.

Consider the parity basis of $\mathcal{R}$. For each $\omega \subseteq X$, let $\omega_i = \omega \cap \{x_{i1}, \cdots, x_{in}\}$ for each i. Then

$$\prod_{x \in \omega}(1 + x) = \prod_{1 \leq i \leq m} \prod_{x_{ij} \in \omega_i}(1 + x_{ij}).$$

If $|\omega_i| > n/2$, we replace $\prod_{x_{ij} \in \omega_i}(1 + x_{ij})$ by

$$\prod_{x_{ij} \in \overline{\omega_i}}(1 + x_{ij}) \cdot \Pi_i = \sum_{S \subseteq [1, m-c]} g_i^S(X) \cdot \prod_{j \in S} h_j(X)$$

in $\mathcal{R}/\mathcal{K}$ (see [12] for details), where $\overline{\omega_i} = \{x_{i1}, \ldots, x_{in}\} - \omega_i$. Note that g_i^S has degree $\leq n/2 + N^\delta$ in $\{x_{i1}, \cdots, x_{in}\}$, and $\leq N^\delta$ for all other x. Thus, we have that

$$\prod_{x \in \omega}(1 + x) = \sum_{S \subseteq [1, m-c]} p_\omega^S(X) \prod_{j \in S} h_j(X),$$

where each $p_\omega^S(X)$ has degree $\leq n/2 + mN^\delta$ in each group of variables $\{x_{i1}, \cdots, x_{in}\}$.

Let M denote the number of monomials in X satisfying this restriction. Then

$$M \leq \Big(\sum_{i=0}^{n/2+mN^\delta} \binom{n}{i}\Big)^m$$

Since $\sum_{i=0}^{n/2}\binom{n}{i} \leq 2^{n-1}$ and $\binom{n}{i} \leq \binom{n}{n/2}$ for each $n/2 < i \leq mN^\delta$, we further get

$$\begin{aligned}
&M \leq \Big(2^{n-1} + \binom{n}{\frac{n}{2}} mN^\delta\Big)^m \\
&\leq \Big(2^{n-1} + \frac{n!mN^\delta}{(\frac{n}{2}!)^2}\Big)^m \\
&\approx 2^{(n-1)m}\Big(1 + \sqrt{\tfrac{8}{\pi n}} mN^\delta\Big)^m, \text{using the Stirling formula } n! \approx \sqrt{2\pi n}(\tfrac{n}{e})^n \\
&\leq 2^{(n-1)m}\Big(1 + \sqrt{\tfrac{8}{\pi}} \tfrac{m^{1+\delta}m}{n^{1/2-\delta}} + o(1)\Big), \text{ since } m^{2+5\delta} \leq n.
\end{aligned}$$

Since there are at most 2^{m-c} terms of the form $\prod_{j \in S} h_j(X)$, and since, for each such term $\prod_{j \in S} h_j(X)$ in $\prod_{x \in \omega}(1 + x)$, there are at most M choices for the associated $p_\omega^S(X)$,

$$dim(\mathcal{R}/\mathcal{K}) \leq 2^{m-c}M \leq \Big(\frac{1}{2^c} + \sqrt{\frac{8}{\pi}}\frac{m^{2+\delta}}{n^{1/2-\delta}2^c} + o(1)\Big)2^N.$$

Let $e(N) = o(1)$. Since $c = \log m + 1$, $dim(\mathcal{R}) + dim(\mathcal{R}/\mathcal{K}) < 2^N$, a contradiction. □

4 Proof of the Theorem

The proof uses a query which encodes the multiple parity problem, similarly to [7].

Proof: Consider the database schema $\{R_1, R_2, R_3\}$, where each R_i is a $(3k+1)$-ary relation. Let *k-Mod-4* be the query defined (for each database instance I of this schema) by

$$k\text{-}Mod\text{-}4(I) = \{(x_1, ..., x_k) \mid N_{x_1,...,x_k}(I) \bmod 4 = 0\}$$

where $N_{x_1,...,x_k}(I) = \Sigma_{i=1}^{3} |\{(y_1, ..., y_{2k+1}) \mid R_i(x_1, ..., x_k, y_1, ..., y_{2k+1})\}|$.

Then *k-Mod-4* $\in$ FOIES$_k$: To maintain the answers to *k-Mod-4*, one just maintains 3 k-ary auxiliary relations S_1, S_2, S_3, where S_i is defined to contain all $(x_1, ..., x_k)$ tuples in D^k such that $N_{x_1,...,x_k}(I)$mod $4 = i$. (It is easy to see that two auxiliary relations, say S_1 and S_2, suffice because they and the query answer allow us to derive S_3.)

To prove that *k-Mod-4* is not in FOIES$_{k-1}$, suppose to the contrary that *k-Mod-4* has a foies $\mathcal{F}$ using at most $(k-1)$-ary auxiliary relations. We first rewrite $\mathcal{F}$ into another foies $\mathcal{F}'$ that uses only $(k-1)$-ary auxiliary relations and suppose $\mathcal{F}'$ has μ auxiliary relations, say $S_1, \ldots, S_\mu$. We pick two constants a and b, a positive integer $\ell > \mu$, and a set D of ℓ constants including a and b.

Let P be a given instance of the multiple parity problem with $m = \ell^k$ and $n = \ell^{2k+1}$. We will show how to solve the multiple parity problem using an AC^0 circuit, through the foies $\mathcal{F}'$ for *k-Mod-4*. Let $t_1, \ldots, t_m$ be an enumeration of D^k and $s_1, \ldots, s_n$ be an enumeration of D^{2k+1}. Let I be the following database instance:

$$I(R_1) = D^k \times \{a\}^{2k+1}$$
$$I(R_2) = I(R_3) = \{(t_i, s_j) \mid x_{i,j} \text{ is } \textit{true} \text{ in the instance } P\}$$

Observe that for each i $(1 \le i \le m)$, if $x_{i1} \oplus x_{i2} \oplus \cdots \oplus x_{in} = 1$ (or 0), then $N_{t_i}(I)$ is 3 mod 4 (or 1 mod 4, respectively). Thus *k-Mod-4*$(I) = \emptyset$.

For each i, $1 \le i \le m$, consider the insertion of the set $\Delta = \{(t_i, b, ..., b)\}$ into R_1. Since $\mathcal{F}'$ is a foies for *k-Mod-4*, there is a first-order query $\phi_i(\Delta, I, I_{\text{aux}})$ which computes the new answer (observe that the current answer is empty). From the definition of the query *k-Mod-4* and the database I, it is clear that the new answer will be either the same as the current answer or the current answer plus the tuple t_i. By building the update Δ into ϕ_i and using the construction of I, we can easily obtain a first-order formula $\varphi_i(I, I_{\text{aux}})$ which returns true iff $x_{i1} \oplus x_{i2} \oplus \cdots \oplus x_{in} = 0$.

We can now construct an AC^0 circuit C to compute the multiple parity problem with $m - \log m - 1$ help bits. We can represent each auxiliary relation S_i by ℓ^{k-1} bits. Therefore, the auxiliary relations $S_1, ..., S_\mu$ can be represented by $\mu \cdot \ell^{k-1} \leq m - \log m - 1$ bits, which will be the help bits. Now for each i, $1 \leq i \leq m$, we construct an AC^0 circuit C_i from φ_i to compute the parity of $x_{i1} \oplus x_{i2} \oplus \cdots \oplus x_{in}$. Since C is an AC^0 circuit, it has a polynomial size (in m). Set δ of Lemma 4 to $\frac{1}{5k}$. Then $n = \ell^{2k+1} = \ell^{k(2+\frac{1}{k})} > \ell^{k(2+\frac{5}{5k})} = m^{2+5\delta}$. Lemma 4 now implies that C must be of size exponential in m, a contradiction. ∎

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases.* Addison-Wesley, 1995.

[2] D. Barrington, A note on the theorem of Razborov, *Notes*(1987).

[3] J.-Y. Cai, Lower bounds for constant-depth circuits in the presence of help bits. *Inform. Proc. Lett.* 36(1990), 79-83. Also appeared in *Proc. of FOCS 89.*

[4] G. Dong and J. Su. Incremental and decremental evaluation of transitive closure queries by first-order queries (extended abstract). In *Proc. 16th Australian Computer Science Conference*, 1993. Full version appears in [5].

[5] G. Dong and J. Su. Incremental and decremental evaluation of transitive closure by first-order queries. *Information and Computation*, 120(1):101–106, July 1995.

[6] G. Dong and J. Su. Space bounded FOIES. In *Proc. of the ACM Symposium on Principles of Database Systems*, pages 139–150, May 1995.

[7] G. Dong and J. Su. Arity Bounds for First-Order Incremental Evaluation and Definition of Polynomial Time Database Queries. *Journal of Computer and System Sciences.* To appear in a special issue for PODS 95.

[8] G. Dong, J. Su, and R. Topor. Nonrecursive incremental evaluation of datalog queries. *Annals of Mathematics and Artificial Intelligence*, 14(2-4):187–223, 1995.

[9] G. Dong and R. Topor. Incremental evaluation of datalog queries. In *Proc. Int'l Conference on Database Theory*, pages 282–296, Berlin, Germany, October 1992. Full version appears as part of [8].

[10] S. Patnaik and N. Immerman. Dyn-FO: A parallel dynamic complexity class. In *Proc. ACM Symp. on Principles of Database Systems*, pages 210–221, 1994.

[11] A. Razborov, Lower bounds for the size of circuits of bounded depth with basis AND, XOR, *Notes*(1987).

[12] R. Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity, *Proc. 19th ACM Symposium on Theory of Comput.*, 77-82, 1987.